Desalination and Water Treatment
www.deswater.com
doi:10.5004/dwt.2017.20795

79 (2017) 161–177
June

Estimating nutrient criteria of the lakes and reservoirs by reference condition approach and stressor-response models

Wenchun Sun[a,b], Shouliang Huo[b,*], Chunzi Ma[b], Beidou Xi[b], Zhuoshi He[b], Jingtian Zhang[b], Jing Su[b], Da An[b]

[a]*School of Government, Beijing Normal University, Beijing 100012, China, email: sunwc@craes.org.cn*
[b]*State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China, Tel. +86-10-84937970, Fax +86-10-84913805, email: huoshouliang@126.com (S. Huo), xiaomachunzi@163.com (C. Ma), xibeidou@263.net (B. Xi), zhuoshihe@163.com (Z. He), wuxiang1998@163.com (J. Zhang), sujing169@163.com (J. Su), anda@craes.org.cn (D. An)*

Received 25 October 2016; Accepted 6 April 2017

ABSTRACT

Accurate and robust approaches for quantifying regional numeric nutrient criteria are critical to the management and restoration of aquatic ecosystems. In this paper, systematic statistical approaches combining reference condition approach and stressor-response models were developed to determine nutrient criteria in Anhui lakes and reservoirs, China. Reference lake method and lake population distribution method served as the reference condition approach were used to identify nutrient criteria by respectively selecting the upper 25th percentile and the lower 25th percentile as the reference condition. The stressor-response models determined by linear regression model (LRM), Bayesian non-hierarchical linear model (BNLM), classification and regression tree (CART), and change point analysis (CPA) were developed to compare and verify the consistency of these methods. Results indicated that there were no significant differences in nutrient criteria determined by the two types of methods. The ranges of numeric nutrient criteria in Anhui lakes and reservoirs were determined as follows: 0.020–0.046 mg/L for TP and 0.42–0.81 mg/L for TN. The advantages, disadvantages, and applicability of each method were discussed and estimated, which would be beneficial in the scientific selection of nutrient criterion approach and improving the feasibility of setting nutrient criteria.

Keywords: Lake and reservoir; Nutrient criterion; Reference condition approach; Stressor-response model

1. Introduction

Eutrophication causing by the excessive input of nitrogen and phosphorus has been threatening the numerous lakes and reservoirs in China. Nutrient criteria have been assumed to be important for regulators to control cultural eutrophication and to protect current and future water quality [1–3]. However, the establishment of numeric nutrient criteria has been proven to be exceedingly difficulty due to intensive human activities and the unavailability of lake watersheds minimally affected [4].

Nutrients such as nitrogen and phosphorus are not toxic to aquatic organisms or humans at low concentrations [5,6], and the dose-response relationships that represent the toxic effects of chemical pollutants using simple laboratory studies have limited applicability to nutrient criteria development. The statistical method based on large amounts of monitoring data would provide the theory and approach foundation for the establishment of nutrient criteria. Three types of scientifically statistical approaches including the reference condition approach, mechanistic modeling, and stressor-response analysis have been widely employed to determine numeric nutrient criteria in US, Europe, and Canada [3,7–9].

*Corresponding author.

1944-3994 / 1944-3986 © 2017 Desalination Publications. All rights reserved.

The reference condition approach, based on the frequency distribution of nutrient data, commonly containing reference lake method and lake population distribution method, is preferred for regions with available reference lakes [10,11]. The Regional Nutrient Criteria Research Plan of China implemented in 2008 preliminarily explored the feasibility of the reference condition approach to support the development of nutrient criteria in China by case studies [12–15]. The mechanistic modeling method requires adequate data to construct the feasible equations for representing waterbodies and to calibrate the parameters in these equations [3]. In addition, the current research on the mechanism of lake eutrophication is faultiness, the mechanistic modeling method is not applied widely [16]. If minimally affected sites cannot be identified and paleoecological or historic data are unavailable in the region of interest, the stressor-response models may be appropriate to set numerical criteria [12,16–18]. Because of the widespread contamination of aquatic ecosystems by industrialization, urbanization, and agriculture, the stressor-response models are more suitable for the establishment of nutrient criteria in China [12].

Recently, linear regression models (LRM) have been developed to explore the stressor-response relationship between nutrient and response variables that are directly or indirectly associated with designated water uses [5,13,14]. For example, a simple linear regression (SLR) model and a multiple linear regression (MLR) model were estimated and interpreted for deriving numeric nutrient criteria to address total nitrogen (TN) and total phosphorus (TP) pollution in the Eastern and Yungui Lake Ecoregion [14,19]. A Bayesian non-hierarchical linear model (BNLM) has been employed to estimate a multilevel linear model for the prediction of chlorophyll a (Chl *a*) from TN and TP [5]. However, the biological response to nutrient gradients might be subtle and difficult to detect with a linear regression analysis [20]. Meanwhile, ecological responses to environmental gradients are often nonlinear, non-normal, and heterogeneous [21]. Hence, non-linear models including a classification and regression tree (CART) and a change point analysis (CPA) are developed for the stressor-response relationship to establish nutrient thresholds [16,22].

In this study, the reference condition approach, including the reference lake method and the lake population distribution method, and the stressor-response models developed by LRM, BNLM, CART, and CPA are compared and analyzed to determine nutrient criteria in Anhui lakes and reservoirs, China. The specific objectives are: 1) to determine nutrient criteria of lakes and reservoirs; 2) to validate the consistency of LRM, BNLM, CART, and CPA results for TN and TP; and 3) to discuss and estimate the advantages, disadvantages, and applicability of these methods for the determination of nutrient criteria.

2. Materials and methods

2.1. Study area

Anhui Province (114°54′-119°37′E, 29°41′-34°38′N) is a monsoon climate zone located in the eastern China. The lakes cover a total area of approximately 1750 km^2, mainly distributed in the watershed of the Yangtze River and Huai River. Some lakes of this area have suffered from serious eutrophication in recent decades and environmental quality continues to decline with the rapid economic development in Anhui Province. In this study, 37 water bodies are investigated and studied to establish nutrient criteria in this region (Fig. 1). The information about the 37 studied lakes and reservoirs are listed in Table S1 (Supporting Information, SI).

2.2. Data sources and data quality

Data for Anhui lakes and reservoirs were obtained from the ambient lake monitoring network, which is supported by the Department of Environmental Protection of Anhui Province. The obtained data consist of measurements for stressor variables such as TN and TP, and response variables such as Chl *a*. A total of 37 water bodies were selected for this analysis, mainly from 1991 to 2013. Data were included for lakes and reservoirs that had at least three surveys every year over this time interval. Six reservoirs minimally impaired by human activities, were identified as reference sites. TN, TP, and Chl *a* were analyzed in laboratory using standard testing procedures as recommended by the Ministry of Environmental Protection of China [23]. The TN was measured by the method of alkaline potassium persulfate digestion with ultraviolet light spectroscopy. The TP was measured by the ammonium molybdate spectrophotometric method. Chl *a* measurements were achieved by the spectrophotometric method.

The detection limits for TP and TN were 0.01 mg/L and 0.1 mg/L, respectively. Observations in the database below the detection limits were assigned values equal to one-half the detection limits because these observations were encountered infrequently (less than 15% of the total dataset). This method of addressing the detection limits has been reported to be sufficiently accurate for determining descriptive statistics such as the mean and standard deviation [11,24,25].

2.3. Methods for setting nutrient criteria

2.3.1. Reference condition approach

The reference condition approach requires judging and discerning reference lakes or sites, and depends on the availability of sufficient data from these reference sites representing the distributions of different variables [3]. The reference lake method and the lake population distribution method, as the reference condition approach, were employed to determine nutrient criteria in Anhui lakes and reservoirs.

Reference sites are relatively undisturbed monitoring points which have minimal human activities and can support all designated water uses [24]. Based on existing and/or new data collected, the upper 25th percentile of the frequency distribution for reference sites can be identified as the nutrient criteria. Generally, the proportion of reference lake could be at least 10% of the lakes and reservoirs per ecoregion [7], cropland and/or urban land ratio is no more than 20% in the watershed, and the reference sites do not connect a drain outlet and/or shoreline directly and there is no obvious endogenous pollution [10,11,26]. Six reservoirs in Anhui Province may meet the above requirements, and are able to serve as the reference lakes minimally impaired by human activities.

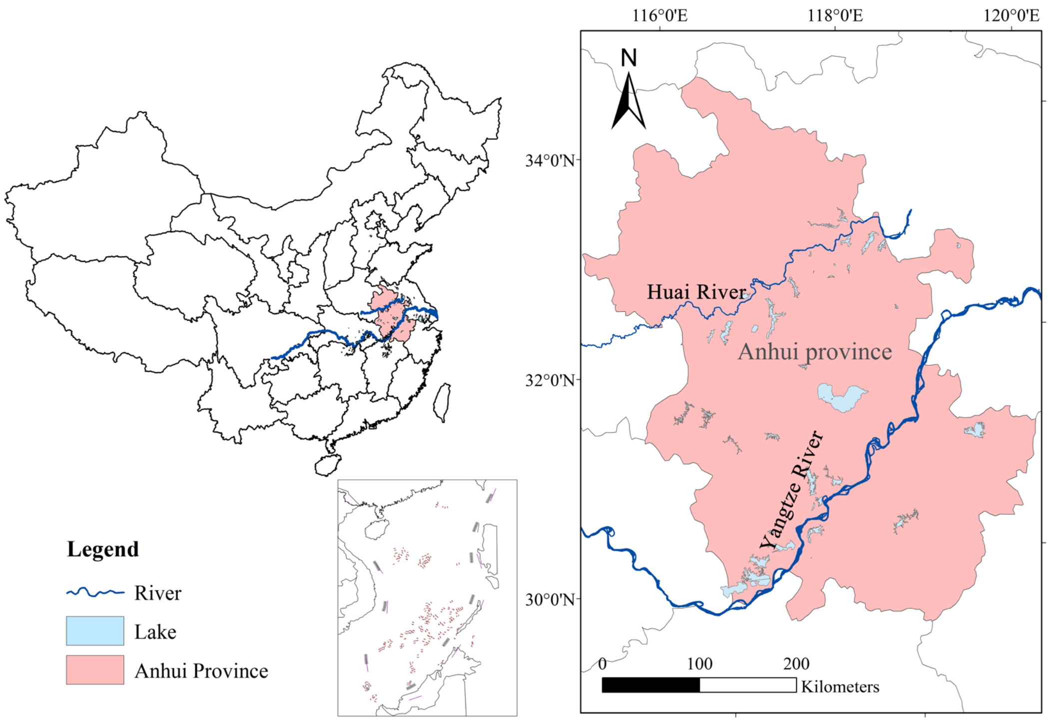


Fig. 1. The locations of studied lakes and reservoirs in Anhui Province.

The population distribution approach does not need to identify the reference lakes or sites; it can set the criterion values by the use of all the lake and reservoir data presently available. The lower 25th percentiles of nutrient reflecting high water quality can be selected as the nutrient criteria. In general, the 25th percentile of the frequency distribution may be insufficient for the protection of lakes and reservoirs if water quality has been severely degraded for most lakes and reservoirs in the region. Therefore, a low percentile may be suggested as the nutrient criteria of this region . If almost all lakes are impacted by human activity to some extent, the 5th percentile is recommended [7].

2.3.2. Stressor-response models

Stressor-response models including LRM, BNLM, CART, and CPA, were used to estimate and interpret for deriving numeric nutrient criteria to address both nitrogen and phosphorus pollution [3,7]. The LRM method provides an estimate of the linear relationship between one response variable and more than one stressor variables such as the concentration of TP and/or TN. The LRM method can be further divided into the simple linear regression (SLR) model and multiple linear regression (MLR) model. The results of SLR are two coefficients specifying the intercept and slope of a straight line representing the modeled relationship between the two variables [3]. MLR is useful in cases in which other environmental factors except nutrients influence the response variable, or in cases in which effects of different nutrients must be modeled together [19].

The BNLM method is able to provide probabilistic predictions, enabling inference at unmonitored sites [27]. The advantages of the BNLM method are 1) the ability of incorporating prior information, 2) the explicit handling of uncertainty, and 3) the straightforward ability to absorb new information [28]. We assume that the response variables and the related parameters for BNLM meet the following distribution in this paper:

$$
\begin{aligned}
& Chla_i \sim N(\mu_i, \sigma^2), \sigma \sim U(0,100), \\
& \log Chla_i = beta0 + beta1 * \log TP_i, or \log Chla_i \\
& \qquad = beta0 + beta1 * \log TN_i, or \\
& \log Chla_i = beta0 + beta1 * \log TP_i + beta2 * \log TN_i \\
& \qquad + beta3 * \log TP_i * \log TN_i, \\
& beta0 \sim N(\mu_{beta0}, \sigma^2_{beta0}), \\
& \mu_{beta0} \sim N(0, 1.0E-6), \sigma_{beta0} \sim U(0,100), \\
& beta1 \sim N(\mu_{beta1}, \sigma^2_{beta1}), \\
& \mu_{beta1} \sim N(0, 1.0E-6), \sigma_{beta1} \sim U(0,100), \\
& beta2 \sim N(\mu_{beta2}, \sigma^2_{beta2}), \\
& \mu_{beta2} \sim N(0, 1.0E-6), \sigma_{beta2} \sim U(0,100), \\
& beta3 \sim N(\mu_{beta3}, \sigma^2_{beta3}), \\
& \mu_{beta3} \sim N(0, 1.0E-6), \sigma_{beta3} \sim U(0,100)
\end{aligned}
\tag{1}
$$

where μ_i is the mean of observation i, σ^2 is the error precision.

The initial values of models are set as: list (sigma = 100, sigma0 = 1, sigma1 = 1, sigma2 = 1, sigma3 = 1, mu.beta0 = 1, mu.beta1 = 1, mu.beta2 = 1, mu.beta3 = 1). In the process of modeling operation, the Gibbs algorithm is used to conduct the simulation analysis by 4000 initial iterations and 26000 iterations to ensure the convergence of parameters.

CART is attractive for exploring environmental and ecological researches due to its capability to address both continuous and discrete variables, predict interactive effects, and establish hierarchical structure [29,30]. CART analysis does not make assumptions about the underlying distribution of the predictor variables. It can accommodate numerical data that are highly skewed or multi-modal as well as categorical predictors. This reduces the time of estimating whether variables are normally distributed and transforming non-normally distributed data [29].

CPA can be realized by nonparametric change point analysis (nCPA) and Bayesian hierarchical modeling (BHM) to establish the nonlinear stressor-response relationship. The CPA is applied to calculate the location of thresholds or change points in bivariate relationships [3]. If observations from multiple sites are ordered along the gradient, a threshold or sudden change in the statistical attributes of the dependent variable will occur in the relationship between a stressor variable and a response variable. CPA can therefore be used to determine the point at which the change occurs [22,31]. In this study, the response variables can be approximated by a normal distribution, and a Gibbs sampling procedure was used to estimate the parameters [32,33]. Before conducting the CPA using the BHM method, specific information on the distribution of the response variable is required to estimate whether the distribution satisfies the assumption of normality.

2.4. Statistical analyses

In this study, TN and TP were chosen to represent the stressor variables, and Chl *a* was selected as the biological response variable. The annual mean data of lakes and reservoirs were applied to build the stressor-response relationship by using LRM, CART, and CPA methods, and the original data were employed by the reference condition approach and the BNLM model. The concentrations of TN, TP, and Chl *a* in Anhui lakes and reservoirs were log transformed (base 10 for the LRM, CART, and CPA; base natural logarithm for the BNLM) to meet the normality assumption [34].

The reference condition approach and LRM analyses were performed by using SPSS 16. The WinBUGS14 software was developed to simulate the BNLM. The R function rpart was used to calculate the node of the CART model and the change point of nCPA, and the R function bootstrap was applied to evaluate the 90% confidence interval of each threshold with 1000 random permutations (R 3.0.2, http://cran.r-project.org/bin/ windows/ base/). Matlab software (R2007b, The MathWorks Company, US) was used for the BHM analysis and the calculation of the 90% confidence interval for the change points.

Reservoirs have the similar characteristics to lakes in nutrients ecological effect and human activities, hence, similar methods can be used to determine reservoir nutrient criteria.

3. Results

3.1. Reference condition established by reference lake method and population distribution method

Six reservoirs being in Anhui Province were selected to serve as the reference sites because of the reservoirs minimally impaired by human activities. Reference values derived from the reference lake method were slightly higher than the values derived from the lake population distribution methods (Table 1). Reference TN concentrations were 0.63 mg/L for the reference lake method and 0.47 mg/L for the lake population distribution method. Reference TP values were 0.028 mg/L and 0.020 mg/L, respectively. Finally, reference Chl *a* had values of 2.8 µg/L and 2.1 µg/L, respectively. This indicated that reference values by using the 25th percentile of the lake population distribution were more conservative than the 75th percentile of reference lakes in this region.

3.2. Stressor-response model for nutrient criteria

3.2.1. Linear regression model

The use of stressor-response models is based on the assumption that lakes and reservoirs within Anhui Province are likely to have similar Chl *a* responses to nutrient variation. The annual values of Chl *a*, TP and TN were collected regularly from this region to build the linear regression models. SLR models of lgChl *a* using lgTP or lgTN as predictor are shown in Fig. 2 and Table 2. Confidence intervals (90%) were used to describe the inherent uncertainty in estimating a mean response value when deriving criteria from stressor-response relationships.

Fig. 2 and Table 2 indicated that there were significant positive correlations ($p < 0.001$) between lgChl *a* and lgTP or lgTN in this region, respectively. The correlation between lgChl *a* and lgTP was much stronger than that between lgChl *a* and lgTN for the region.

To accurately predict future conditions using regression models, residual values against predicted values are plotted in Fig. S1 (SI). It could be seen that the scatter of the residual values was constant over the entire range of fitted values and randomly distributed around zero. This suggested that sampling variance was constant, and certain inferences from the regression relationship may be accurate. The distributions of the error in observed values of the dependent variable about the estimated relationship are shown in Fig. S2 (SI). As shown in Fig. S2 (SI), quantile-quantile plots provided a robust, graphical approach for assessing whether residuals were normally distributed [34]. Most values clustered around the solid line, indicating a near-normal distribution. Departures of samples at the upper and lower end from a straight line suggested that the residuals extended to slightly more extreme values than predicted by a normal distribution.

The prediction intervals and confidence intervals can provide useful information when deriving nutrient criteria

Table 1
Statistical results by reference lake method and lake population distribution method

Method	Variable	Percentage, %							N
		5	15	25	50	75	85	95	
Reference lake method	TN (mg/L)	0.25	0.32	0.36	0.46	0.63	0.75	1.01	2217
	TP (mg/L)	0.010	0.012	0.013	0.019	0.028	0.036	0.056	2213
	Chl *a* (μg/L)	1.2	1.6	2.0	2.4	2.8	3.2	6.0	1976
Lake population distribution method	TN (mg/L)	0.30	0.39	0.47	0.84	1.7	2.4	4.6	5288
	TP (mg/L)	0.011	0.015	0.020	0.047	0.13	0.21	0.42	5550
	Chl *a* (μg/L)	1.0	1.7	2.1	2.9	8.1	16.0	39.8	4190

Note: bold text-reference condition.

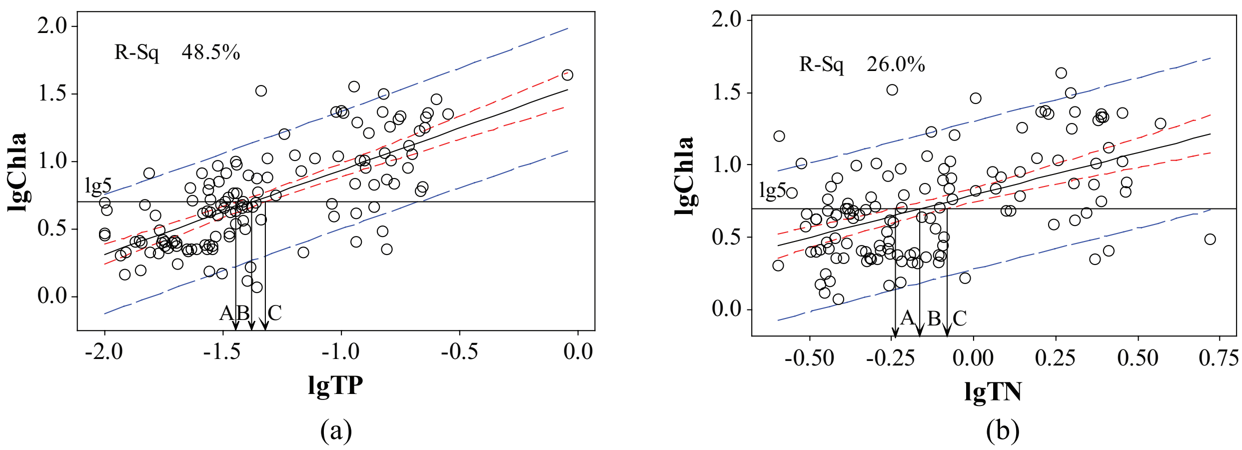


Fig. 2. The simple linear regression models for Anhui lakes and reservoirs. (a) the relationship between lgTP and lgChl *a*; (b) the relationship between lgTN and lgChl *a*.

Table 2
Results of SLR and MLR for the lakes and reservoirs in Anhui Province (TP and TN unit: mg/L)

Variables	Coefficient	*p*	R^2	N	Predicted variable	Predicted value	90% confidence interval	
							Lower limit	Upper limit
Intercept (b)	1.56	<0.001**	0.485	128	TP	0.041	0.036	0.048
lgTP	0.622	<0.001**						
Intercept (b)	0.792	<0.001**	0.260	126	TN	0.69	0.57	0.83
lgTN	0.585	<0.001**						
Intercept (b)	1.492	<0.001**	0.485	126	–	–	–	–
lgTP	0.567	<0.001**			–	–	–	–
lgTN	0.090	0.370			–	–	–	–

Note: ** correlations are significant at $P < 0.01$ (two-tailed).

from SLR [3]. Compared with using prediction intervals to derive a range of possible criteria, criteria associated with confidence intervals spanned a narrower range because that the confidence interval depicts the inherent uncertainty of predicted values. The hypothesis that 5 μg/L Chl *a* concentration was served as the criteria of the response variable for lakes and reservoirs in this region to satisfy the drinking water use [19]. The upper 90% confidence interval intersected Chl *a* = 5.0 μg/L at TP = 0.036 mg/L, TN = 0.57 mg/L, and the lower confidence interval intersected at TP = 0.048 mg/L, TN = 0.83 mg/L, and the mean relationship intersected at TP = 0.041 mg/L, TN = 0.69 mg/L (Arrows A, C, and B, respectively in Fig. 2 (a) and (b)), respectively. The confidence intervals of SLM for TP and TN in Anhui lakes and reservoirs were 0.036–0.048 mg/L and 0.57–0.83 mg/L, respectively.

Furthermore, lgTN and lgTP were simultaneously used to construct a multiple regression model for lgChl *a* in

Anhui lakes and reservoirs. This model was not effective and un-useful for accurately predicting future conditions in this region because the *p*-value of TN was much more than 0.05 (Table 2).

3.2.2. Bayesian non-hierarchical linear model

Taking into account the characteristics of data, the original data of 37 lakes and reservoirs were integrated to determine the stressor-response relationship using the BNLM. Based on the assumed condition, the simple and multivariate BNLMs were estimated for the prediction of lnChl *a* using predictors lnTP or/and lnTN. The results of the BNLM for lnChl *a*-lnTP, lnChl *a*-lnTN, and lnChl *a*-lnTP + lnTN in this region are listed in Table 3. As shown in Table 3, the obtained MC errors were lower than the 10% of standard deviation (SD), which indicating that *a posteriori* estimation was accurate through the preliminary judgment.

The diagram of kernel density estimation can be developed to reflect the variation tendency of mean values and confidence intervals for beta0, beta1, beta2, and beta3 (Fig. S3, SI). Moreover, the autocorrelation function was analyzed to diagnose whether the estimated values of model could reach the goal of convergence.

It can be seen that the kernel density of lnChl *a*-lnTP, lnChl *a*-lnTN, and lnChl *a*-lnTP+lnTN models approximately satisfied the normal distribution (Fig. S3, SI), manifesting that the estimated values of BNLM for lnChl *a*-lnTP, lnChl *a*-lnTN, and lnChl *a*-lnTP + lnTN reached the simulation requirements by WinBUGS software. The autocorrelation functions of lnChl *a*-lnTP and lnChl *a*-lnTN models soon falled to zero with the increase of iterations, showing that the iteration processes had converged (Fig. S4, SI). While the decreasing tendencies of autocorrelation functions for lnChl *a*-lnTP + lnTN model were not significant, only the parameter beta3 had faster convergence speed. This indicated that only the related parameters for lnChl *a*-lnTP and lnChl *a*-lnTN can well satisfy the prior distribution information of BNLM, and can be employed to build the BNLM.

The BNLMs for lnChl *a*-lnTP and lnChl *a*-lnTN were built using the original data in this region, and the corresponding equations were shown as following:

$$\begin{cases} \ln Chl\ a = 4.431 + 0.792 * \ln TP \\ \ln Chl\ a = 2.302 + 0.806 * \ln TN \end{cases} \tag{2}$$

According to the above equations, the deduced TP and TN criteria of 0.028 mg/L and 0.42 mg/L were required to maintain an average Chl *a* concentration of 5 μg/L. The correlations between lnChl *a* and lnTP/lnTN were low, which indicated there were non-linear relationship existed between stressor variables and response variables for the original data. Hence, the non-linear models (such as CART and CPA) were employed for the further research.

3.2.3. Classification and regression tree model

CART can be acted as a variable selection method to identify important factors associated with the variation of the response variable. Chl *a* was used as the response variable for the CART model. The selected variable firstly meets the requirements which is the most important one or has the greatest influence on the Chl *a* concentration. TP and TN observations were included as potential predictor variables. The final models were selected based on their predictability, which was simulated by cross-validation. The CART analyses indicated that a hierarchical structure existed between nutrients and Chl *a* (Fig. 3). The standard error (SE) of the Chl *a* concentration data was used as a measure of dispersion.

The variability of Chl *a* in this region was driven primarily by TP (Fig. 3). The mean Chl *a* concentration for TP less than 0.045 mg/L was 3.47 μg/L (with a standard deviation of 1.68), and the mean Chl *a* concentration at higher TP concentrations was 10.8 μg/L (2.14). For TP lower than 0.045 mg/L, TN was the next most important variable. The other splits were all made on TN. The lower panel of Fig. 3 shows boxplots of Chl *a* concentrations within each of the terminal nodes. Only the first five splits including in the terminal model demonstrated that further splits would not reduce the model's relative predictive error or increase the predictive correlation coefficient [29].

The CART models presented in this study might not be developed to predict Chl *a* concentrations. However, the model can provide valuable information for water quality management. For example, the boxplots show that a large variance in the Chl *a* concentration corresponds to high TP and high TN concentrations. The variations in TP and TN concentrations in this region must be controlled simultaneously, to effectively improve water quality.

Table 3
The results of the BNLM for lnChl *a*-lnTP, lnChl *a*-lnTN, and lnChl *a*-lnTP + lnTN (SD-Standard deviation)

Variable	Parameter	Mean	SD	MC error	2.50%	Median	97.50%	Initial iteration	Iteration
TP	beta0	4.431	0.042	0.001	4.348	4.432	4.511	4001	26000
	beta1	0.792	0.025	0.001	0.742	0.791	0.843	4001	26000
TN	beta0	2.302	0.041	0.001	2.219	2.303	2.382	4001	26000
	beta1	0.806	0.030	0.001	0.746	0.806	0.865	4001	26000
TP + TN	beta0	4.567	0.094	0.006	4.378	4.570	4.746	4001	26000
	beta1	0.908	0.051	0.003	0.810	0.907	1.009	4001	26000
	beta2	−0.202	0.071	0.004	−0.343	−0.199	−0.071	4001	26000
	beta3	−0.253	0.045	0.002	−0.344	−0.251	−0.167	4001	26000

Note: MC error = $SD / \sqrt{n}$

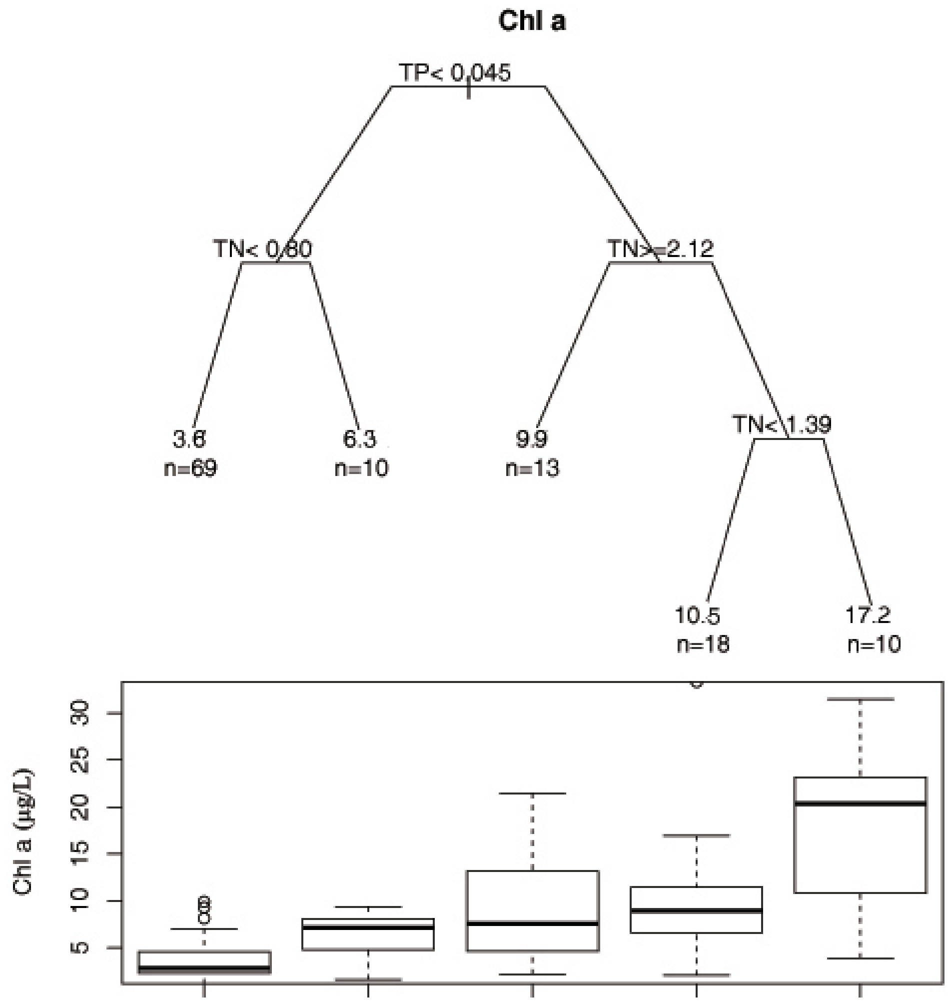


Fig. 3. Regression tree plot of observed Chl *a* partitioned with TN and TP concentrations. The boxplots which represent the Chl *a* concentrations in each terminal node, are ordered from small to large by the mean of Chl *a* concentrations in categorical data.

3.2.4. Changepoint analysis

The CART results for each node in the regression tree were verified by the CPA method. The changepoints, the mean and standard deviation (SD) of the response variable Chl *a* on both sides of the changepoints were estimated using the nCPA and the BHM methods (Table 3). Uncertainty in the changepoint location can be quantified using the range of the middle 90% of the 1000 bootstrap simulation replicates for the nCPA method and the 90% confidence intervals for the BHM method.

The distributions of TP and TN, along with the Chl *a* data based on the annual data, are illustrated in Fig. 4. The results from the nCPA method were comparable to those generated from the BHM method. There were no significant differences between the changepoints from the nCPA and BHM methods for TP and TN, indicating that the probability distribution assumptions for the response variable under the BHM method were appropriate. Because the BHM method utilized the distributional information for the response variable, it generated narrower confidence intervals for the changepoint (see Fig. 4) [22]. If the true probability distribution of the response variable cannot be determined, the nCPA method should be used to confirm the changepoint. Table 5 lists the advantages, disadvantages and applicability of the stressor-response models.

The changepoint analysis was conducted by ordering observations along a stressor gradient (x axis of TP or TN) and identifying the point along the gradient that divides the response variable into the two groups with the greatest difference in deviation. Accordingly, the nCPA method split the dataset into two groups around each unique value of the stressor variable, and then calculated the difference between the deviation for the entire dataset and the sum of the deviations of the two groups. The changepoint was defined as the point that maximizes this difference. Abrupt changes in the response variables were observed, ranging from 0.045 mg/L to 0.046 mg/L for TP and 0.81 mg/L for TN using the nCPA and BHM methods (Table 3).

3.3. Establishment of nutrient criteria

The results obtained by the reference lake method, the lake population distribution method, LRM, BNLM, CART, and CPA in Anhui lakes and reservoirs are listed in Table 6.

Table 4
Threshold values for TP and TN concentrations (mg/L) with Chl *a* concentrations (μg/L) determined using the nCPA and BHM methods (standard deviation: SD)

Method		TP	TN
nCPA	Changepoint	0.045	0.81
	Confidence interval	0.042, 0.094	0.72, 1.34
	Chl *a* mean [n] ± SD	3.5 [79] ± 1.68, 10.8. [50] ± 2.14	3.7 [81] ± 1.87, 9.8 [461] ± 2.13
BHM	Changepoint	0.046	0.81
	Confidence interval	0.044, 0.046	0.80, 0.84
	Chl *a* mean [n] ± SD	3.5 [80] ± 1.68, 11.1 [49] ± 2.12	3.7 [81] ± 1.87, 9.8 [461] ± 2.13

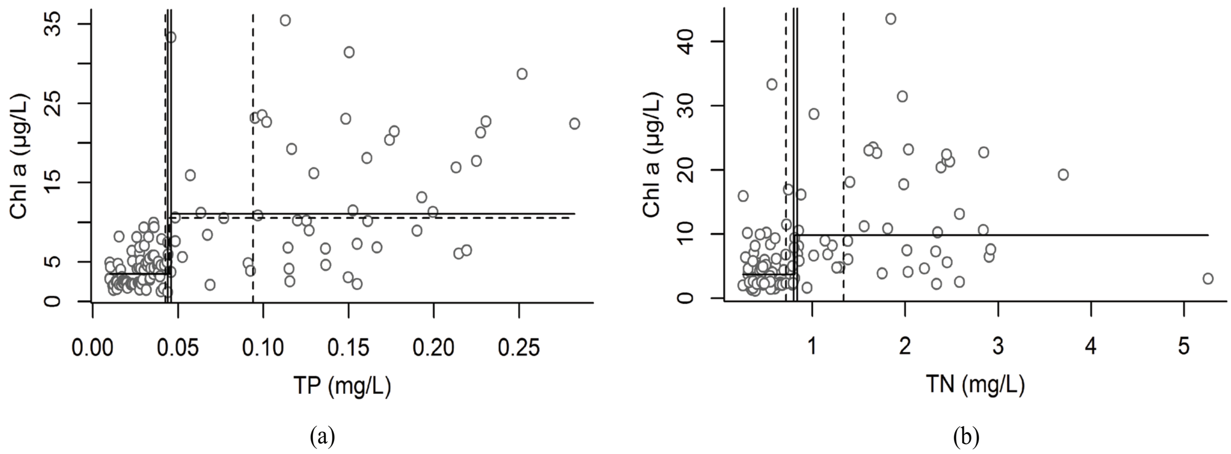


Fig. 4 Changepoint distributions of (a) TP and (b) TN estimated for Chl *a* using the nCPA (dashed lines) and BHM (solid lines) methods. Polyline shows modeled response, with a step increase at changepoint. Vertical lines show the 90% confidence intervals about the changepoint. Data are shown as open circles.

As shown in Table 6, the nutrient threshold values determined by the CART method were nearly bracketed by the values derived from the nCPA and BHM methods in Anhui Province. The criteria obtained by the reference lake method, the lake population distribution method, LRM and BNLM were obviously lower than those obtained by CART, nCPA and BHM methods. A Friedman test suggested that there were no significant differences in the results obtained from the various statistical methods ($p > 0.05$). Hence, the range of nutrient criteria in Anhui Province could be defined as: 0.020–0.046 mg/L for TP and 0.42–0.81 mg/L for TN. The nutrient criterion ranges by the lake population distribution approach and the LRM method were 0.014–0.043 mg/L and 0.020–0.032 mg/L for TP, 0.36–0.78 mg/L and 0.25–0.47 mg/L for TN, respectively, in the Eastern Plain Ecoregion (including North, Mid-East, and Southeast Lake Ecoregion) [14,16,19,35], which were in good consistency with the scope of nutrient criteria in Anhui lakes and reservoirs. In the USA, nutrient criteria of ecoregions V (South Central Cultivated Great Plains) and IX (Southern Temperate Forested Plains and Hills), which are located in the same climate zone as Anhui Province, were similar to nutrient criteria in Anhui lakes and reservoirs (see Table 6) [5, 38]. In the case of Chl *a* less than 5.0 μg/L, the probability of TN less than 0.81 mg/L is 71.2%, and the probability of TP lower than 0.046 mg/L is 74.1%. This indicated that the criteria values obtained by the various methods were scientific and reasonable for Anhui lakes and reservoirs.

4. Discussion

A systematic method, including the reference lake method, the lake population distribution method, LRM, BNLM, CART, and CPA, was developed to determine nutrient criteria in Anhui lakes and reservoirs. The reference conditions were unavailable if the aquatic ecosystems are widely contaminated by industrialization, urbanization, and agriculture in the region of interest. In this case, changepoints or thresholds were considered as reference concentrations to determine nutrient criteria. Hence, reference conditions were compared with changepoints and thresholds in this paper. The nutrient criteria determined by the various methods were approximate and unanimous, and at least two types of potential approaches were available to establish criterion values which offered some degree of certainty. Nutrient criterion thresholds or ranges determined by the different methods in various countries and regions are listed in Table 6. There were no significant differences in the results obtained by the various statisti-

Table 5
The advantages, disadvantages and applicability of the stressor-response models

Model	Advantage	Disadvantage	Applicability
LRM	1) Quantitatively analyses the influence of nutrient increase on the response variables; 2) not require identifying the reference lakes or least impacted lakes in the studied region, and not require collecting amount of history data.	1) Need to set the response criteria; 2) the data need to be classified to eliminate the influence of confounding factors on the relationship; 3) the model extrapolation would introduce greater uncertainty; and 4) the land use-nutrient regression model could not quantify all sources of anthropogenic influence.	Heavy impacted by human activities and good linear relationship between stressor variable and response variable existed in the region.
BNLM	1) Adjust the influence of covariate variables on the all levels to predict for the variability of the output result; and 2) effectively relieve the data missing and measuring error to evaluate and compare the relative heterogeneity, and to avoid under or over estimates.	1) Need to set the response criteria; 2) the modeling dataset are generated randomly, which leads to slightly different in the calculated result.	
CART	1) A non-parametric method and makes no assumptions about the underlying distribution of values of the predictor variables; and 2) not require the establishment of a response threshold value.	Model is lack of robustness when the number of samples is small, and the accuracy of the results might not be guaranteed.	Heavy impacted by human activities existed, ecological responses to environmental gradients are nonlinear, non-normal, and heterogeneous, and the stressor-response relationship cannot be expressed by the linear models in the region.
CPA	1) Evaluate the positions of thresholds or changepoints in binary relationships and provide natural candidates for nutrient criteria; and 2) not require the establishment of a response threshold value.	1) Additional analyses are required to determine whether the characteristics of the chosen value are consistent with a protective target; and 2) require estimating whether the values of the response variable at values below the changepoint support the designated uses of waterbodies.	

cal methods for the same region. For example, the nutrient criteria determined by the CART, nCPA, and BHM methods were almost consistent for the seven lake ecoregions, China, and were in the ranges of criteria obtained by the LRM method (except Yungui Lake Ecoregion) [16,19]. For the Yungui Lake Ecoregion, the criteria determined by the reference condition method were in consistency with the values derived from the LRM method [19,36]. Table 6 also showed there were significant regional differences in the nutrient criteria among various regions and indicated that, in the absence of human activities, environmental factors rather than TP and TN (e.g., salinity, light, temperature, water color, and suspended sediment) would promote or inhibit the growth of algae [16].

All the methods could be used simultaneously to provide various criteria for comparison. Due to some ecoregions lack a large amount of data, an order of preference to apply these difference techniques should be considered to setting criterion values. Where minimally impacted reference lakes exist in a region, the reference lake method should be given preference to determine the regional nutrient criteria [7,11,36]. Where undisturbed or nearly undisturbed conditions are difficult to identify, but sufficient data are available in a region, the population distribution method could be preferred to determine the regional nutrient criteria [35]. If degraded conditions prevail and appropriate data exist to adequately quantify the relationship of variables, the stressor-response relationship would be preferred to provide a statistically defensible method [5,11]. If only one method can be used to determine criteria due to the limitation of data quantity and type, professional judgment and expertise will be required to be implemented [13].

The reference condition method is to determine nutrient criteria based on the frequency distribution of the original survey data. It was identified as one of the most straightforward methods for setting criteria because the data includes natural variability [39]. However, the selection of percentage would easily introduce subjective bias, and the attainment of designated water uses was not considered by the reference condition approach for determining nutrient criteria. The scientific determination of reference lakes or reference sites also had restricted the application of the reference condition method. Lakes or sites impacted minimally by human disturbance are scarce in developed regions and historic data are few and even problematic [40]. The establishment of criterion thresholds based on percentiles might be skewed by a bias in the data toward either pristine or highly influenced sites, which would introduce great uncertainty. Hence, more research should be developed to quantify sources of uncertainty, to account for the uncertainty and the correlative impact of uncertainty [40].

The stressor-response relationship between nutrient and Chl *a* was considered as the best method to develop nutrient criteria for the widespread contamination of aquatic ecosystems by industrialization, urbanization, and agriculture [16]. Compared to the reference condition approach,

Table 6
Nutrient criterion thresholds or ranges determined by the different methods in various countries and regions

Country or region		Variables	Reference condition approach			Stressor-response models					References
			Reference lake method	lake population distribution approach	Trisection method	LRM	BNLM	CART	nCPA	BHM	
China	Anhui Province	TN (mg/L)	0.63	0.47	–	0.69	0.42	0.80	0.81	0.81	This study
		TP (mg/L)	0.028	0.02	–	0.042	0.028	0.045	0.045	0.046	
	Eastern Plain Ecoregion	TN (mg/L)	–	0.360–0.785	0.67	0.250–0.474	–	–	–	–	[14,35]
		TP (mg/L)	–	0.014–0.043	0.029	0.020–0.032	–	–	–	–	
	North Lake Ecoregion	TN (mg/L)	–	–	–	0.846 ± 0.473	–	1.685	0.708	0.698	[16,19,36]
		TP (mg/L)	–	–	–	0.019 ± 0.0039	–	0.0224	0.0224	0.0224	
	Mid–East Lake Ecoregion	TN (mg/L)	–	–	–	0.374 ± 0.139	–	0.643	0.579	0.565	
		TP (mg/L)	–	–	–	0.0218 ± 0.0074	–	0.0381	0.0381	0.038	
	Southeast Lake Ecoregion	TN (mg/L)	–	0.31–0.65	0.55	–	–	0.532	0.511	0.51	
		TP (mg/L)	–	0.010–0.025	0.02	0.0293 ± 0.01	–	0.0298	0.0298	0.025	
	Northeast Lake Ecoregion	TN (mg/L)	–	–	–	0.863 ± 0.64	–	0.643	0.67	0.67	
		TP (mg/L)	–	–	–	0.053 ± 0.021	–	0.0255	0.0255	0.022	
	Ganxin Lake Ecoregion	TN (mg/L)	–	–	–	0.421 ± 0.293	–	0.508	0.877	0.885	
		TP (mg/L)	–	–	–	0.016 ± 0.005	–	0.0473	0.0473	0.0482	
	Ningmeng Lake Ecoregion	TN (mg/L)	–	–	–	–	–	1.745	1.745	1.75	
		TP (mg/L)	–	–	–	–	–	0.146	0.159	0.157	
	Yungui Lake Ecoregion	TN (mg/L)	0.175	0.37	0.21	0.173 ± 0.049	–	0.494	0.494	0.3	
		TP (mg/L)	0.01	0.01	0.01	0.008 ± 0.0018	–	0.0185	0.0154	0.015	
Brazil	São Paulo State	TN (mg/L)	–	–	0.35	–	–	–	–	–	[37]
		TP (mg/L)	–	–	0.01	–	–	–	–	–	
USA	Ecoregion V	TN (mg/L)	–	0.56	–	–	–	–	–	–	[5,38]
		TP (mg/L)	–	0.033	–	–	–	–	–	–	
	Ecoregion IX	TN (mg/L)	–	0.36	–	–	–	–	–	–	
		TP (mg/L)	–	0.02	–	–	–	–	–	–	

the stressor-response models do not require identifying reference lakes and obtaining a large amount of data from reference or minimally impacted lakes in the studied region. The majority lakes of China are suffering from significant human activities, and no or minimally impacted reference lakes are usually not available in the ecoregions. Hence, stressor-response models are more suitable for determining nutrient criteria.

The stressor-response model could be classified as the linear and the non-linear relationship. The LRM and BNLM methods would be employed to establish the linear regression relationship between stressor and response variables. They could be able to quantify the influence of response variables with the increase of stressor variables. In the region that good linear relationship existed between nutrient and Chl *a*, LRM and BNLM can be used to deduce reliable nutrient criteria. The BNLM method combines prior data and the actual monitoring data. For the small sample data, it can reduce error and *a priori* data, and nearly had no influence on the calculation result for the great sample data. The application of prior information and the Markov chain Monte Carlo (MCMC) simulation method in BNLM can effectively relieve the data missing and measuring error to evaluate and compare the relative heterogeneity, and to avoid under or over estimates. The BNLM provided a new pathway and method to correctly clarify and explain the relationship between nutrients and Chl *a*.

However, the linear relationship also has disadvantages, which hampered the application for the special-type lakes to some extent. 1) The stressor-response relationship between nutrients and algae was susceptible to being confounded with environment factors, such as turbidity, water temperature, light, and lake area. Hence, these confounding factors should be identified or included in future models [12,19,41]. 2) The biological response to nutrient gradients might be subtle, nonlinear, non-normal, and heterogeneous, which would likely be difficult to detect with a linear regression analysis [20,21]. 3) If there are many high anthropogenic impact lakes in a specific ecoregion, the predicted data would be required far from the data points to extrapolate the relationship [3,19]. 4) The linear stressor-response relationship requires the establishment of a threshold value for the response variable to determine a potential numerical criterion, which would introduce some subjective biases.

The non-linear relationship combining CART and CPA methods could be developed to verify the validation of the linear results. Compared with the linear stressor-response relationships, the non-linear models do not require the setting of a response threshold value to determine numerical criterion [3]. A threshold or a changepoint refers to the positions of stressor variable leading to abrupt changes in both the mean and the variance of the ecological response variable, which provides natural candidates for nutrient criteria [22,31]. The CART and CPA methods can be used to explore the subtle non-linear relationship between the stressor and response variables, and do not require the hypothesis of classical regression, such as independence, normality, linear or smoothness [16]. The CPA could potentially provide a criterion to determine this threshold value; additional analyses are required to estimate whether the values of the response variable at values below the changepoint would support the designated uses of waterbodies.

The CART method can be used to analyze the influence of various factors (such as nutrients, environment, and lake type) on the response variable Chl *a* and to explore the significance of factors on the response variable under various concentrations [28]. As the CPA method, the nCPA and BHM methods estimate the relationship between a response variable and a stressor variable to identify the changepoint. The impact of other factors on the response variable was not considered in the nCPA and BHM analysis; thus, the threshold value may incorporate greater uncertainty. The nCPA method does not make probabilistic assumptions about the response data; it is more robust, and the related calculations are more straightforward than the BHM method. However, the nCPA method could not effectively use the information about the probabilistic distribution of the response variable, which is less efficient than the BHM method when such information exists [22].

The CART analysis recommended a hierarchical structure for Chl *a*, TN, and TP, and the thresholds in the regression tree model were nearly consistent with those observed when assessed individually by the nCPA and BHM methods. The information derived from the hierarchical structure might be useful in the establishment of nutrient criteria. Therefore, utilizing hierarchical structure as a tool to understand large datasets may stimulate the development of nutrient criteria [20].

Although more and more datasets on hydrology, chemistry, and biology of lakes and reservoirs could be obtained from national water quality monitoring networks, the lack of sufficient data was still the biggest obstacle for the development of a water quality management plan. The complexity of natural processes in lakes and reservoirs made it difficult to transform routine monitoring data into scientific knowledge that can be developed to support lake-specific management decision. The integration of the reference condition approach and stressor-response model would be beneficial to determine scientific and reasonable nutrient thresholds in lakes and reservoirs.

5. Conclusions

Developing accurate and effective ways to estimate numeric nutrient criteria are critical for the management and recovery of aquatic ecosystems. The reference condition approach and stressor-response models were synthesized to set nutrient criteria, which proposes an attempt in supporting policies for eutrophication control, and provides a reference for establishing nutrient criteria. The ranges of numerical nutrient criteria were 0.020–0.046 mg/L for TP and 0.42–0.81 mg/L for TN, which may control the growth of algae in Anhui lakes and reservoirs. The various methods determined nutrient criteria should be comprehensively considered when providing technical support and implementing water quality standards for regulation to avoid the under or over protection of lakes and reservoirs. In addition, the process described will be benefit to support countries and regions with similar climate characteristics in incorporating reference condition approach and stressor-response models into their numeric criteria development programs and to achieve the further water management goal of reducing nitrogen/phosphorus pollution.

Acknowledgement

This study is supported by the Mega-projects of Science Research for Water Environment Improvement (Program No. 2012ZX07101-002) and the National Natural Science Foundation of China (No. 41521003).

Supporting information available

Fig. S1: Residuals from regression fit plotted versus predicted lgChl *a*: (a) TP; (b) TN. Fig. S2: Quantile-quantile plot comparing residuals from the relationship: (a) TP; (b) TN. Fig. S3: The diagram of kernel density estimation for (a) lnChl *a*-lnTP, (b) lnChl *a*-lnTN, and (c) lnChl *a*-lnTP + lnTN. Fig. S4: The autocorrelation function for (a) lnChl *a*-lnTP, (b) lnChl *a*-lnTN, and (c) lnChl *a*-lnTP + lnTN. Table S1: The information about the 37 studied lakes and reservoirs.

References

[1] W.K. Dodds, E.B. Welch, Establishing nutrient criteria in streams, J. N. Am. Benthol. Soc., 19 (2000) 186–196.
[2] P.C. Hawkins, J.R. Olson, R.A. Hill, The reference condition: predicting benchmarks for ecological and water-quality assessments, J. N. Am. Benthol. Soc., 29 (2010) 312–343.
[3] U.S. EPA, Using Stressor-response Relationships to Derive Numeric Nutrient Criteria, EPA-820-S-10-001. U.S. Environmental Protection Agency, Office of Water, Washington DC, 2010.
[4] W.M. Lewis, Yield of nitrogen from minimally disturbed watersheds of the United States, Biogeochemistry, 57/58 (2002) 375–385.
[5] E.C. Lamon, S.S. Qian, Regional scale stressor-response models in aquatic ecosystems, J. Am. Water Resour. As., 44 (2008) 771–781.
[6] C.A. Stow, Y.K. Cha, S.S. Qian, A Bayesian hierarchical model to guide development and evaluation of substance objectives under the 2012 Great Lakes Water Quality Agreement, J. Great Lakes Res., 40 (Supplement 3) (2014) 49–55.
[7] U.S. EPA, Nutrient criteria technical guidance manual: lakes and reservoirs[M], EPA-822-B-00-001. United States Environment Protection Agency, Office of Water, Washington, DC, 2000.
[8] M. Ramin, S. Stremilov, T. Labencki, A. Gudimov, D. Boyd, G.B. Arhonditsis, Integration of numerical modeling and Bayesian analysis for setting water quality criteria in Hamilton Harbour, Ontario, Canada, Environ. Modell. Softw., 26 (2011) 337–353.
[9] G. Bouleau, D. Pont, Did you say reference conditions? Ecological and socio-economic perspectives on the European Water Framework Directive, Environ. Sci. Policy, 47 (2015) 32–41.
[10] W.K. Dodds, R.M. Oakes, A technique for establishing reference nutrient concentrations across watersheds affected by humans, Limnol. Oceanogr.-Meth., 2 (2004) 333–341.
[11] W.K. Dodds, Determining ecoregional reference conditions for nutrients, secchi depth and chlorophyll a in Kansas lakes and reservoirs, Lake Reserv. Manage., 22 (2006) 151–159.
[12] S.L. Huo, B.D. Xi, C.Z. Ma, H.L. Liu, Stressor−response models: a practical application for the development of lake nutrient criteria in China, Environ. Sci. Technol., 47 (2013) 11922–11923.
[13] S.L. Huo, C.Z. Ma, B.D. Xi, Z.H. Tong, Z.S. He, J. Su, F.C. Wu, Determining ecoregional numeric nutrient criteria by stressor-response models in Yungui ecoregion lakes, China, Environ. Sci. Pollut. R., 21 (2014) 8831–8846.
[14] Y.L. Zhang, S.L. Huo,·C.Z. Ma, B.D. Xi, X. Li, H.L. Liu, Using stressor–response models to derive numeric nutrient criteria for lakes in the eastern plain ecoregion, China, CLEAN-Soil Air Water, 42 (2014) 1509–1517.
[15] J.B. Chen, J. Lu, Establishment of reference conditions for nutrients in an intensive agricultural watershed, Eastern, Environ. Sci. Pollut. R., 21 (2014) 2496–2505.
[16] S.L. Huo, C.Z. Ma., B.D. Xi, Z.S. He, J. Su, F.C. Wu, Nonparametric approaches for estimating regional lake nutrient thresholds, Ecol. Indic., 58 (2015) 225–234.
[17] M.F. Bowman, K.M. Somers, Considerations when using the reference condition approach for bioassessment of freshwater ecosystems, Water Qual. Res. J. Can., 40 (2005) 347–360.
[18] J.L. Stoddard, D.P. Larsen, C.P. Hawkins, R.K. Johnson, R.H. Norris, Setting expectations for ecological condition of running waters: the concept of reference condition, Ecol. Appl., 16 (2006) 1267–1276.
[19] S.L. Huo, C.Z. Ma, B.D. Xi, R.T. Gao, X.Z. Deng, T.T. Jiang, Z.S. He, J. Su, F.C. Wu, H.L. Liu, Lake ecoregions and nutrient criteria development in China, Ecol. Indic., 4 (2014) 61–10.
[20] E.H. Brian, T.J. Scott, D.L. Scott, Sestonic chlorophyll- *a* shows hierarchical structure and thresholds with nutrients across the Red River basin, USA, J. Environ. Qual., 42 (2013) 437–445.
[21] P. Legendre, L. Legendre, Numerical Ecology, 2nd ed, Elsevier, Amsterdam, The Netherlands, 1998.
[22] S.S. Qian, R.S. King, C.J. Richardson, Two methods for the detection of environmental thresholds, Ecol. Model., 166 (2003) 87–97.
[23] PRC MEP, Water and wastewater monitoring and analysis methods, China Environmental Science Press, Beijing, 2002. (In Chinese).
[24] M.W. Suplee, A. Varghese, J. Cleland, Developing nutrient criteria for streams: an evaluation of the frequency distribution method, J. Am. Water Resour. As., 43 (2007) 453–472.
[25] US EPA, Data Quality Assessment: Statistical Methods for Practitioners, US Environmental Protection Agency, Office of Water, Washington, DC, 2006.
[26] M. Wallin, T. Wiederholm, R.K. Johnson, Guidance on Establishing Reference Conditions and Ecological Status Class Boundaries for Inland Surface Waters, Final Version 7.0, The Swedish Environmental Protection Agency, 2003.
[27] D.G. Clayton, Generalized Linear Mixed Models. In: Markov Chain Monte Carlo in Practice, Chapman and Hall, London, UK, 1996.
[28] O. Malve, S.S. Qian, Estimating nutrients and Chlorophyll a relationships in Finnish Lakes, Environ. Sci. Technol., 40 (2006) 7848–7853.
[29] S.S. Qian, C.W. Anderson, Exploring factors controlling the variability of pesticide concentrations in the willamette river basin using treebased models, Environ. Sci. Technol., 33 (1999) 3332–3340.
[30] S.S. Qian, Environmental and Ecological Statistics with R, Chapman & Hall/CRC, London, UK, 2009.
[31] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, Wadsworth Statistics/Probability, Chapman & Hall/CRC, London, UK, 1984.
[32] M. Lavielle, E. Lebarbier, An application of MCMC methods for the multiple change-points problem, Signal Process., 81 (2001) 39–53.
[33] S.S. Qian, Y.D. Pan, R.S. King, Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data, Ecol. Indic., 4 (2004) 29–37.
[34] A.C. Tamhane, D.D. Dunlop, Statistics and data analysis from Elementary to Intermediate, Prentice Hall, New Jersey, United States 2000.
[35] S.L. Huo, B.D. Xi, J. Su, F.Y. Zan, Q. Chen, D.F. Ji, C.Z. Ma, Determining reference conditions for TN, TP, SD and Chl-a in eastern plain ecoregion lakes, China, J. Environ. Sci., 25 (2013) 1001–1006.

[36] S.L. Huo, F.Y. Zan, Q. Chen, B.D. Xi, J. Su, D.F. Ji, Q.G. Xu, Determining reference conditions for nutrients, chlorophyll a and Secchi depth in Yungui Plateau ecoregion lakes, China, Water Environ. J., 26 (2012) 324–334.

[37] D.G.F. Cunha, A.P. Ogura, M.D.C. Calijuri, Nutrient reference concentrations and trophic state boundaries in subtropical reservoirs, Water Sci. Technol., (2012) 1461–1467.

[38] U.S. EPA, Ecoregional Nutrient Criteria Documents for Lakes & Reservoirs, 2001, http://www2.epa.gov/nutrient-policy-data/ecoregional-nutrient-criteria-documents-lakes-reservoirs.

[39] A.L. Solheim, Reference Conditions of European Lakes: Indicators and methods for the Water Framework Directive Assessment of Reference conditions, Draft Version 5; Dec. 30, 2005.

[40] P.A. Chambers, D.J. McGoldrick, R. Brua, G.A. Benoy, Development of environmental thresholds for Nitrogen and Phosphorus in streams, J. Environ. Qual., 41 (2012) 7–20.

[41] S.L. Wong, J. Chang, Salinity and light effects on growth, photosynthesis, and respiration of *Grateloupia filicina (Rhodophyta)*, Aquaculture, 182 (2000) 387–395.

Supporting information

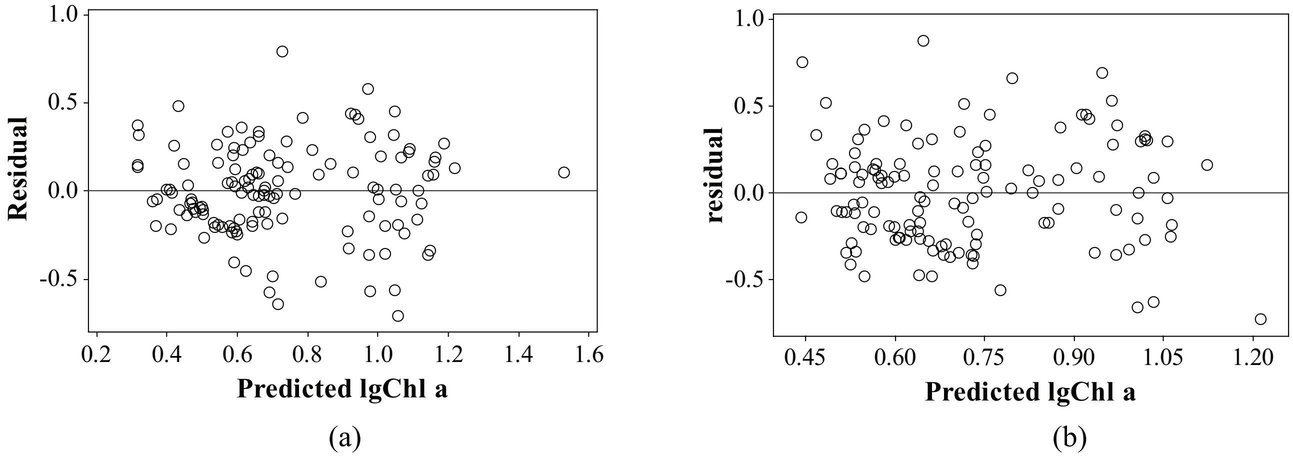


Fig. S1. Residuals from regression fit plotted versus predicted lgChl *a*: (a) TP; (b) TN.

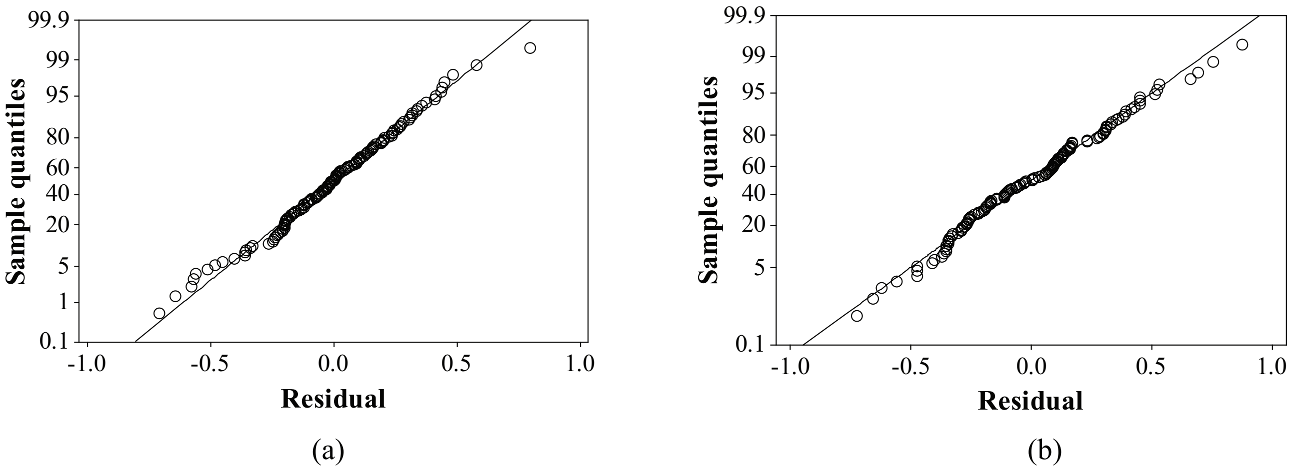
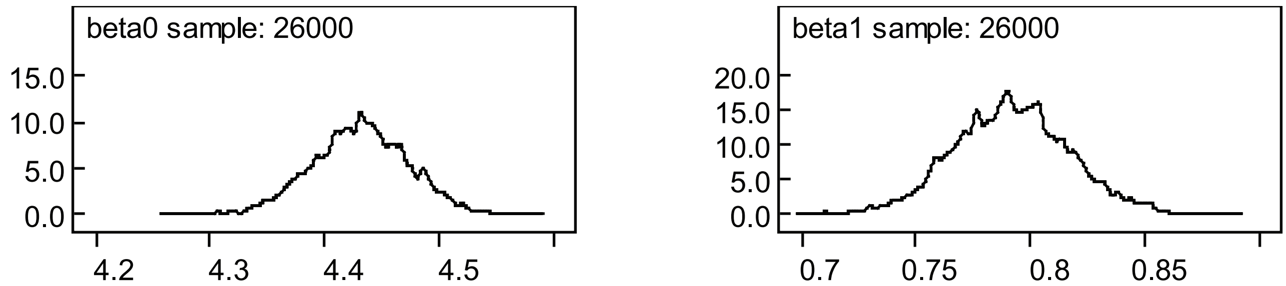
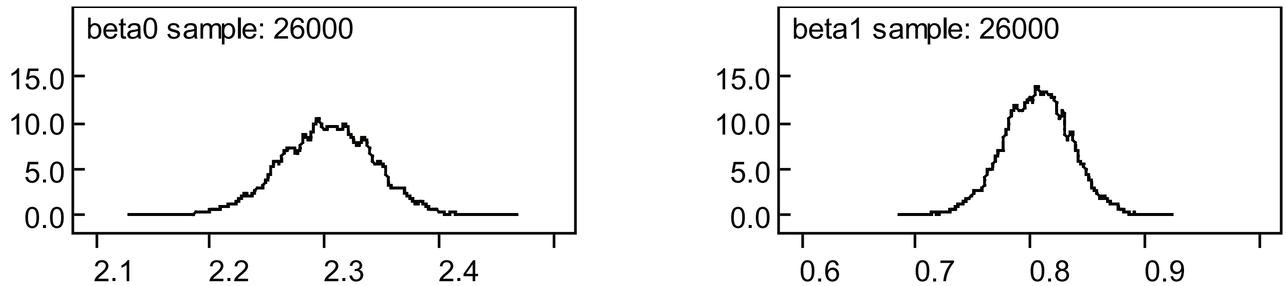


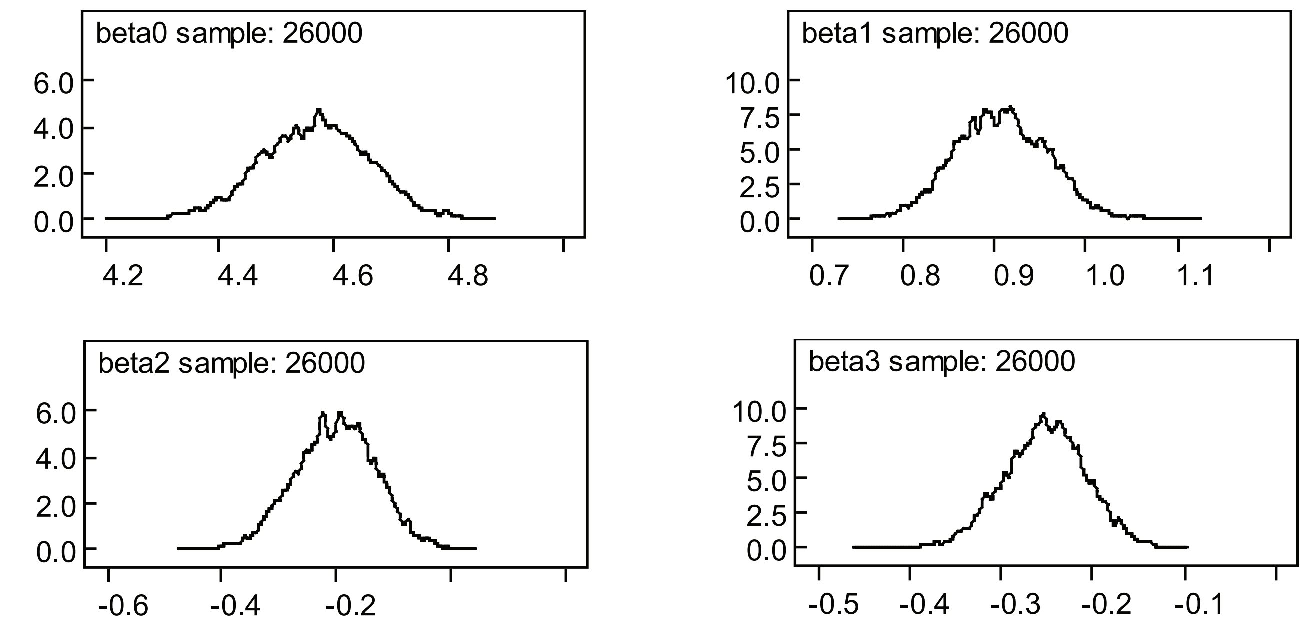
Fig. S2. Quantile-quantile plot comparing residuals from the relationship: (a) TP; (b) TN.



(a) lnChl *a*-lnTP

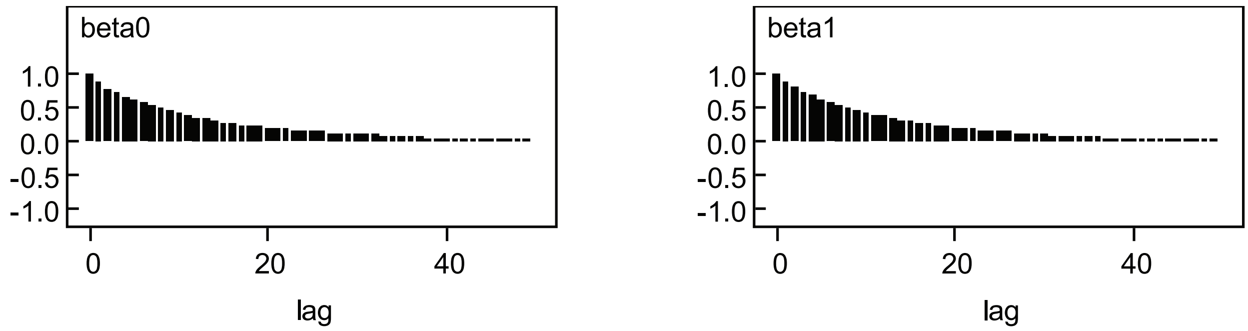


(b) lnChl *a*-lnTN

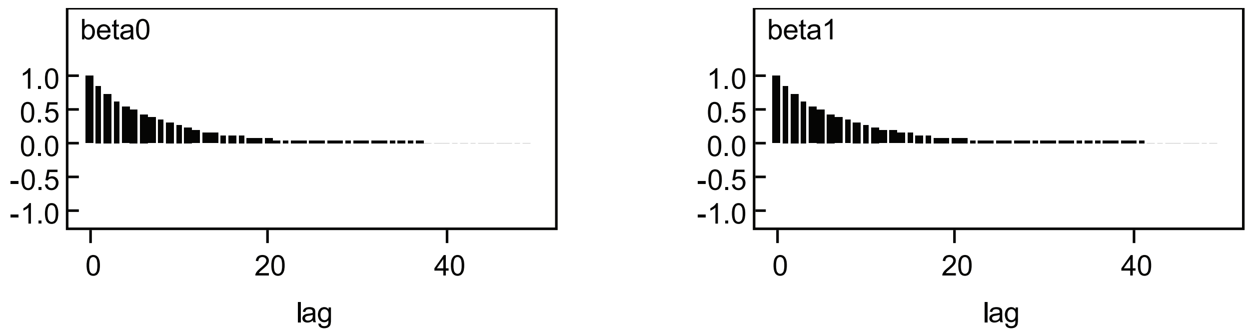


(c) lnChl *a*-lnTP+lnTN

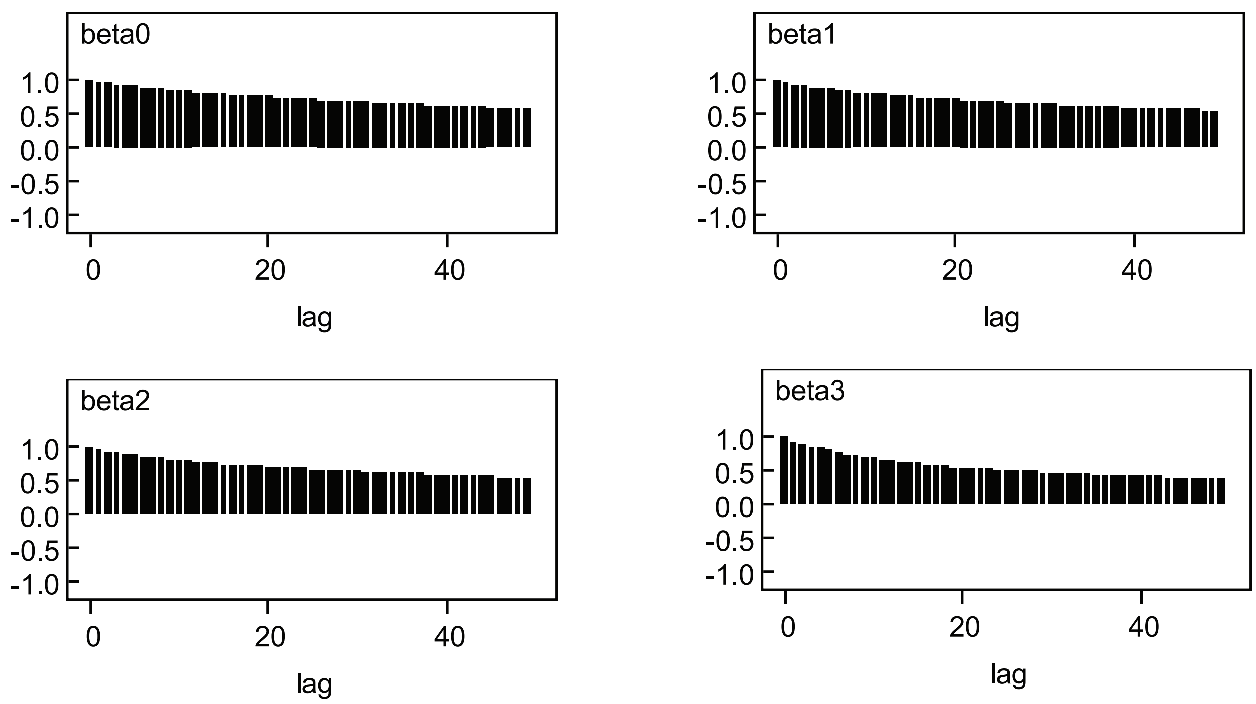
Fig. S3. The diagram of kernel density estimation for (a) lnChl *a*-lnTP, (b) lnChl *a*-lnTN, and (c) lnChl *a*-lnTP + lnTN.



(a) lnChl *a*-lnTP



(b) lnChl *a*-lnTN



(c) lnChl *a*- lnTP+lnTN

Fig. S4. The autocorrelation function for (a) lnChl *a*-lnTP, (b) lnChl *a*-lnTN, and (c) lnChl *a*-lnTP+lnTN.

Table S1
The information about the 37 studied lakes and reservoirs

NO.	Lake Name	Position (Latitude and Longitude)	Lake Area (Km^2)	Mean/ Maximum Depth (m)	Water Volume (× 10^8 m^3)	Catchment area (km^2)	Age (y)	Type	Reference Lake YES or NO
1	Anfeng Tang	116°38´-116°44´E, 32°16´-32°20´N	36.42	2.67/3.60	0.97	390	~2600	Reservoir	NO
2	Baidang Lake	117°19´-117°27´E, 30°47´-30°51´N	39.67	3.06/4.50	1.21	775	–	Lake	NO
3	Bo Lake	116°19´-116°33´E, 30°04´-30°15´N	180.40	4.41/6.86	7.94	1087	~1600	Lake	NO
4	Caizi Lake	117°01´-117°09´E, 30°43´-30°58´N	172.10	1.67/8.28	2.87	3234	–	Lake	NO
5	Chao Lake	117°16´-117°05´E, 31°25´-31°43´N	769.55	2.69/3.77	20.70	9258	~12000	Lake	NO
6	Chengdong Lake	116°18´-116°28´E, 32°12´-32°22´N	120.00	1.50/2.60	2.10	2128	~2.60 × 10^6	Lake	NO
7	Chengxi Lake	116°01´-116°18´E, 32°11´-32°33´N	199.00	2.70/3.90	5.37	1750	~2.60 × 10^6	Lake	NO
8	Chengxi Reservoir	118°16´-118°18´E, 32°19´-32°21´N	–	–	0.85	168	60	Reservoir	NO
9	Dongpu Reservoir	117°08´-117°12´E, 31°52´-31°55´N	–	–	2.42	207.5	61	Reservoir	NO
10	Fengyangshan Reservoir	117°33´-117°38´E, 32°40´-32°42´N	–	–	10.35	146	59	Reservoir	NO
11	Foziling Reservoir	116°13´-116°20´E, 31°15´-31°21´N	–	–	4.96	1840	65	Reservoir	YES
12	Gaotang Lake	117°07´-117°13´E, 32°34´-32°44´N	49.00	1.73/2.50	0.85	400	–	Lake	NO
13	Guangou Reservoir	117°23′-117°25′E, 32°43′-32°45′ N	–	–	–	92	–	Reservoir	NO
14	Huangda Lake	116°14′-116°33′E, 29°56′-30°08′ N	299.20	3.94/5.30	11.79	1686	~2200	Lake	NO
15	Huayuan Lake	117°45´-117°53´E, 32°55´-33°02´N	34.00	1.35/2.10	0.50	875	~2×10^8	Lake	NO
16	Jiaogang Lake	116°34´-116°41´E, 32°15´-32°18´N	40.00	0.44/1.1	0.18	480	–	Lake	NO
17	Longgan Lake	115°19´-116°17´E, 29°52´-30°05´N	316.20	3. 78/4.58	11.96	5511	~690	Lake	NO
18	Longhekou Reservoir	116°39´-116°48´E, 31°15´-31°20´N	50.00	–	8.20	1111	59	Reservoir	YES
19	Longzi Lake	117°23´-117°25´E, 32°53´-32°56´N	7.80	–	–	143	–	Lake	NO
20	Lutang Reservoir	117°39´-117°40´E, 32°47´-32°48´N	2.28	–	0.14	33.6	–	Reservoir	NO
21	Matang Lake	116°34´-116°38´E, 30°24´-30°27´N	8.50	–	0.40	87	–	Lake	NO
22	Meishan Reservoir	115°42´-115°55´E, 31°30´-31°42´N	–	–	23.37	2100	63	Reservoir	YES
23	Mozitan Reservoir	116°20´-116°23´E, 31°11´-31°14´N	–	–	3.39	570	61	Reservoir	YES
24	Nanyi Lake	118°50´-119°02´E, 31°03´-31°10´N	148.40	2.25/3.25	3.34	3369	~1400	Lake	NO
25	Nvshan Lake	117°58´-118°14´E, 32°50´-33°02´N	104.60	1.71/2.40	1.78	4215	~7.3 × 10^8	Lake	NO
26	Pogang Lake	117°04´-117°13´E, 30°33´-30°42´N	60.00	1.52/2.50	0.91	346	–	Lake	NO
27	Qili Lake	118°09´-118°15´E, 32°51´-32°57´N	37.71	–	–	889	~2 × 10^8	Lake	NO
28	Randeng Reservoir	117°47´-117°49´E, 32°43´-32°46´N	9.27	–	–	173	–	Reservoir	NO
29	Shengjin Lake	116°58´-117°14´E, 30°15´-30°28´N	78.48	1.26/3.50	0.99	1554	~3 × 10^6	Lake	NO
30	Shitang Lake	117°04´-117°07´E, 30°36´-30°40´N	11.00	–	–	97	–	Lake	NO
31	Taiping Lake	117°38´-118°26´E, 29°58´-30°29´N	88.00	–	–	2800	59.00	Reservoir	YES
32	Tuo Lake	117°45´-117°51´E, 33°09´-33°17´N	40.00	1.20/2.00	0.48	2983	–	Lake	NO
33	Wabu Lake	116°48´-117°01´E, 32°23´-32°33´N	163.00	2.42/4.15	3.94	800	–	Lake	NO
34	Wuchang Lake	116°36´-116°53´E, 30°14´-30°20´N	100.50	3.43/4.31	3.45	1084	–	Lake	NO
35	Xianghongdian Reservoir	115°59´-116°09´E, 31°26´-31°37´N	–	–	–	1400	61.00	Reservoir	YES
36	Xiangjian Lake	117°15´-117°45´E, 33°06´-33°12´N	45.00	0.93/4.89	0.42	8173	–	Lake	NO
37	Yi Lake	118°59´-119°04´E, 32°46´-32°50´N	18.00	1.70/2.30	0.31	245	–	Lake	NO